REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 2002	Technical Report

4. TITLE AND SUBTITLE
ENVIRONMENTAL MEDICINE GENOME BANK (EMGB): HARDY-WEINBERG EQUILIBRIUM AT AN EOTAXIN LOCUS ON CHROMOSOME 17

5. FUNDING NUMBERS

6. AUTHOR(S)

SPC Murad Dikeidek, SPC Amy Bastille, LTC Larry A. Sonna, M.D., Ph.D., Craig M. Lilly, M.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Research Institute of Environmental Medicine
Kansas Street
Natick, MA 01760-5007

8. PERFORMING ORGANIZATION REPORT NUMBER

T-02/14

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Same as Block 7.

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The Environmental Medicine Genome Bank (EMGB) was used to test a low cost, high throughput, polymerase chain reaction (PCR) - based genetic strategy to distinguish 3 genotypes for a single nucleotide polymorphism (SNP) in the eotaxin gene located on chromosome 17. Using amplification refractory mutati on system PCR (ARMS-PCR), we determined the eotaxin ALA23-THR23 genotypes of 233 samples in the EMGB. The observed allele frequencies were then used to determine the distribution of genotypes that would be expected from the assumptions of the Hardy-Weinberg equilibrium. It was found that, for the overall cohort and for all but one small, heterogeneous subpopulation, the EMGB was in Hardy-Weinberg equilibrium at this locus. The EMGB can therefore serve as a useful source of control material for studies of genes located near this locus at cytogenetic position 17q21.

14. SUBJECT TERMS

Eotaxin, Hardy-Weinberg equilibrium, genetics, chromosome 17, asthma

15. NUMBER OF PAGES
15

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

USARIEM TECHNICAL REPORT T02/14

ENVIRONMENTAL MEDICINE GENOME BANK (EMGB): HARDY-WEINBERG EQUILIBRIUM AT AN EOTAXIN LOCUS ON CHROMOSOME 17

SPC Murad Dikeidek[1], SPC Amy Bastille[2], LTC Larry A. Sonna, MD, PhD[2,3], and Craig M.Lilly, MD[3]

[1]Military Nutrition Division
[2]Thermal and Mountain Medicine Division
U.S. Army Research Institute of Environmental Medicine
Natick, MA 01760

[3]Division of Pulmonary and Critical Care Medicine
Brigham and Women's Hospital
75 Francis St.
Boston, MA 02115

January 2002

US ARMY RESEARCH INSTITUTE OF ENVIRONMENTAL MEDICINE
Natick, MA 01760

DISCLAIMERS

The views, opinions and/or assertions contained in this publication are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

Human subjects participated in these studies after giving their free and informed voluntary consent. Investigators adhered to AR70-25 and USAMRMC Regulation 70-25 on the Use of Volunteers in Research. For protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law CFR 46.

Approved public release; distribution unlimited.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Environmental Medicine Genome Bank (EMGB) was used to test a low cost, high throughput, polymerase chain reaction (PCR) - based genetic strategy to distinguish 3 genotypes for a single nucleotide polymorphism (SNP) in the eotaxin gene located on chromosome 17. Using amplification refractory mutation system PCR (ARMS-PCR), we determined the eotaxin ALA_{23}-THR_{23} genotypes of 233 samples in the EMGB. The observed allele frequencies were then used to determine the distribution of genotypes that would be expected from the assumptions of the Hardy-Weinberg equilibrium. It was found that, for the overall cohort and for all but one small, heterogeneous subpopulation, the EMGB was in Hardy-Weinberg equilibrium at this locus. The EMGB can therefore serve as a useful source of control material for studies of genes located near this locus at cytogenetic position 17q21.

INTRODUCTION

Certain genes may contribute to specific aspects of human performance (1) and to environmental illness. The US Army Research Institute of Environmental Medicine (USARIEM) Environmental Medicine Genome Bank (EMGB) is an ongoing effort to identify genes that correlate with environmental injuries and illnesses and with human physical performance (11, 13). To accomplish this, the EMGB banks DNA samples from human volunteers who have participated in USARIEM environmental and human performance studies, and maintains a registry of phenotypic information. An accurate, low cost, high throughput genetic strategy that can distinguish 3 different genotypes (homozygous wild-type, heterozygous, homozygous variant) of a specific genetic marker, such as the single nucleotide polymorphism (SNP) in the eotaxin gene, can be used to characterize the population of the EMGB. By evaluating the frequency of a genetic marker of known chromosome location, it can be determined whether the EMGB is in Hardy-Weinberg equilibrium, i.e., that the distribution of genotypes at that locus does not deviate significantly from the distribution that one would predict from the measured allele frequencies ($= p^2 + 2\,pq + q^2$, where p and q represent allele frequencies and p^2, pq and q^2 represent the occurrence in the population of homozygous wild-type, heterozygous and homozygous mutant genotypes, respectively). Populations that are in Hardy-Weinberg equilibrium at a particular locus are generally said not to be under a selection pressure at that locus (5). Deviations from Hardy-Weinberg equilibrium, by contrast, can occur under a number of conditions, including natural selection, migration, random drift, mutation (5). Additionally, deviations from Hardy-Weinberg equilibrium can occur in the presence of population stratification (i.e., the existence of distinguishable subpopulations within the main population under study, such as racial subgroups). A theoretical example is presented in Appendix A.

The eotaxin gene is located on chromosome 17 at cytogenetic position 17q21.1 – q21.2 (4). It codes for a chemokine, present in the epithelial cells of airways, that activates a chemokine receptor (called the CCR_3 receptor) that is present on allergy-associated cells, including eosinophils and basophils (7). The CCR_3 chemokine receptor functions in the process that results in the extravasation of eosinophils to tissues of the lung and skin. Past studies have investigated correlations of plasma eotaxin levels with asthma severity. These studies have demonstrated a direct relationship of increased plasma eotaxin levels with increased asthma severity (6). These results in turn have lead to further investigation of eotaxin and its genetic sequence.

An SNP in the eotaxin gene sequence has been associated in vitro with impaired eotaxin secretion in stably transfected human 293 cells (8). This variant was termed THR_{23} due to a substitution of a threonine residue for an alanine at the terminal amino acid of the peptide leader sequence in the wild-type gene (8). Studies have found that the substitution of a polar residue, such as threonine, for a nonpolar residue, such as alanine, decreases the efficiency of signal

peptidases (2, 3, 10). In principle, such a decrease in signal peptidase efficiency might explain the observed impairment in eotaxin secretion. Furthermore, because eotaxin is a chemoattractant for eosinophils, one might predict that a decreased ability to secrete eotaxin would lead to decreased recruitment of eosinophils to an inflamed area. In this regard, it is noteworthy that a case-control study of 119 individuals has demonstrated that THR_{23} homozygotes (n=17) had lower plasma levels of eotaxin and eosinophil counts than homozygous wild-type subjects; heterozygotes had intermediate levels (6, 8).

We set out to determine the eotaxin ALA_{23}-THR_{23} genotype of 233 subjects in the Environmental Medicine Genome Bank. Because the EMGB contains samples from subjects of widely diverse geographic and ethnic backgrounds (representing at least 44 different US states and all major US ethnic subgroups as of July 2000, (13)) our hypothesis was that the EMGB samples would be in Hardy-Weinberg equilibrium at this locus on chromosome 17.

METHODS

SAMPLES

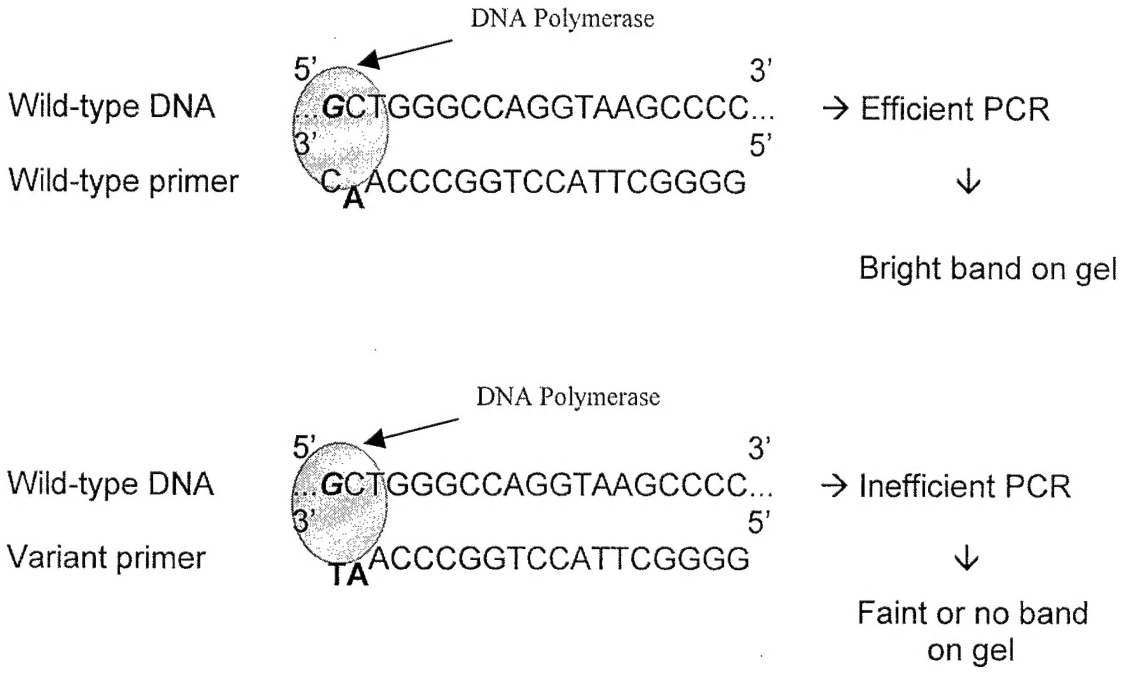
Samples were obtained from the USARIEM Environmental Medicine Genome Bank (EMGB), the composition of which is described in detail elsewhere (11, 13). An attempt was made to obtain a genotype on all 235 samples available in the bank at the time of the study (November 1999 – January 2000) .

ARMS PCR AND GENOTYPE ASSIGNMENTS

Amplification refractory mutation system polymerase chain reaction (ARMS PCR) is a method to determine genotypes that arise from single nucleotide polymorphisms (9). In contrast with traditional PCR methods, which use primers that are identically matched to the sequences under study, ARMS PCR takes advantage of fact that the polymerase chain reaction becomes progressively inefficient as base pair mismatches are introduced into the primer sequences. Therefore, if a primer is placed in a mixture of genomic DNA containing a sequence to which it is mismatched at a single base pair and a sequence to which it is mismatched to at two contiguous base pairs, it will preferentially initiate replication of the better-matched sequence, leading to selective amplification of that allele (Figure 1). It is therefore possible to design primers that will selectively amplify either wild-type or variant sequences. Importantly, since each person carries two alleles of any given gene (one inherited from the father and one from the mother), determining a genotype requires two PCR reactions: One that preferentially amplifies wild-type alleles and one that preferentially amplifies variant alleles (Figures 1 and 2). If only the wild-type primer produces a PCR product, the individual is homozygous wild type. If only the variant primer produces a PCR product, the individual is homozygous variant. If both primers result in a PCR product, the individual is heterozygous.

Figure 1. Theoretical basis of ARMS-PCR. The SNP in the genomic DNA is designated in bold font, as are the mismatches in the primer sequences

Reactions with the Wild-Type allele



Reactions with the Variant allele

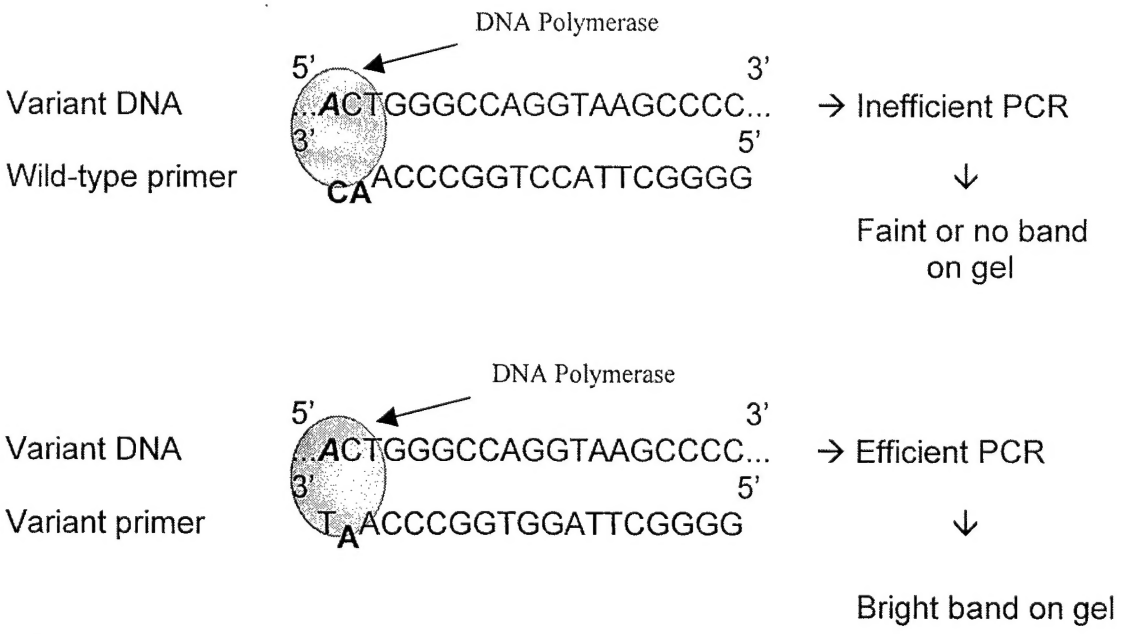
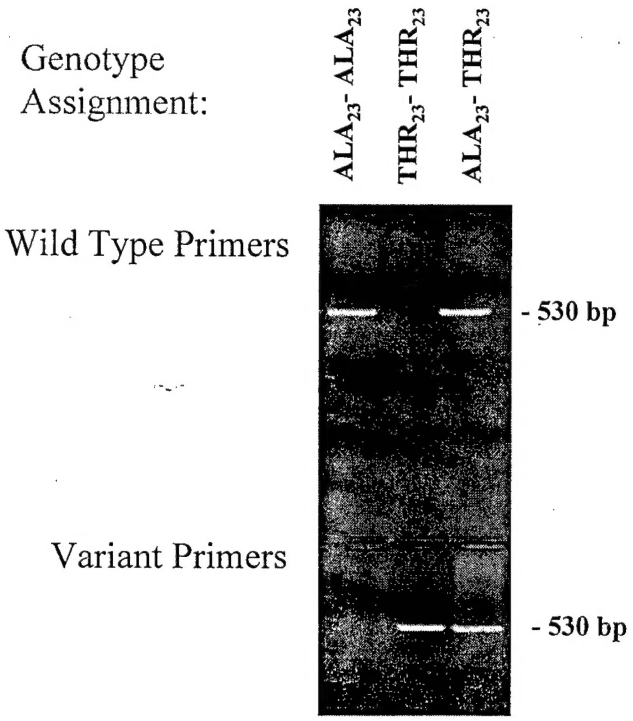


Figure 2. Assignment of eoxtaxin genotypes by ARMS PCR. The assigned genotypes are noted at the top of the gel (ALA_{23}-ALA_{23} = 'wild-type'; THR_{23}-THR_{23} = 'variant'; ALA_{23}-THR_{23} = 'heterozygous')



DNA samples (100 ng each) from the EMGB were placed into the wells of 96-well plates. The final concentrations of reactants, in a total volume of 20 μl (after hot-start addition of Taq, as described below), were: 0.1 μM sense primer (Research Genetics), 0.1 μM anti-sense primer (Research Genetics), 0.2 mM deoxynucleotide triphosphates (dNTP) mix (i.e., 0.2 mM each of dATP, dCTP, dGTP and dTTP) (Boehringer Manneheim), Mg-free PCR buffer (Promega), 2 mM $MgCl_2$ (Promega), and 0.25 U of Taq polymerase (Promega). In this study, the wild-type and variant primers were designed as antisense primers. The wild-type primer sequence was 5'-GGGGCTTACCTGGCCCA**AC**-3', and the variant primer sequence was 5'-GGGGCTTACCTGGCCCA**AT**-3'. The sequence of the sense primer, which binds to a sequence located away from the point mutation and is thus identical in both 'wild-type' and 'variant' PCRs, was 5'-TCAAGGAAGGTTCTTAGATCG-3'.

A hot start method was used in both the wild-type and variant PCRs, in which Taq polymerase is added to the mixture during the first DNA denaturation step (while the samples are still at 94°C). In both reactions, samples were denatured at 94°C for 10 minutes, with Taq polymerase added at 5 minutes. The samples were then subjected to 40 cycles of PCR, with annealing temperatures

of 56°C (for the wild-type reactions) or 53°C (for the variant reactions) for 30 seconds, extension temperatures of 72°C for 1 minute, and denaturing temperatures of 96°C for 40 seconds. The cycling process was followed by a incubation at a final extension temperature of 72°C for 5 minutes.

The electrophoretic mobility of each resulting PCR product was analyzed using a 2% agarose gel (Ultra Pure). The gel was run for one hour at 100 volts. A 100 bp ladder (Promega) was run concurrently to aid in the identification of the amplified DNA bands (= 530 base pairs).

PCR reactions (wild-type and variant) were run in tandem on each batch of samples. The products of the wild-type and variant reactions were placed on the same gel for ease of assigning genotypes. Each PCR run was accompanied by a simultaneous run on five controls of known genotype (previously determined both by direct cycle sequencing and single stranded conformational polymorphism (SSCP) analysis). An allele was said to be present in a sample if the PCR product band for the reaction identifying that allele was at least as bright as the respective control band. Each EMGB sample was genotyped at least twice; if the two genotype assignments were not consistent, the sample was repeated until a consensus genotype could be assigned. Three independent observers assigned genotypes.

HARDY-WEINBERG EQUILIBRIUM CALCULATION

The expected distribution of eotaxin genotypes ($p^2 + 2pq + q^2$) was calculated from the measured allele frequencies. The resulting distribution was then compared to the observed distribution by Chi-square, using only a single degree of freedom (although there are three possible genotypes, they result from combinations of two alleles; since all non-*p* alleles must be *q*, the entire distribution can be computed knowing only the value of *p*, thus only one degree of freedom exists). A *P* value of 0.05 or less was taken to mean that a significant difference exists between the observed distribution and the distribution expected from the assumptions underlying the Hardy-Weinberg equilibrium (principally, that mating occurs at random and there is no selection pressure for one or another allele). Because mating does not in fact occur at random in the US population, a Hardy-Weinberg *P* value was computed for each ethnic subgroup as well as for the entire population studied.

RESULTS

Of 235 samples tested, consensus genotypes were obtained on 233. In the remaining two samples, insufficient sample was present to perform the genotyping reactions twice. Table 1 lists the demographic characteristics of the sample donors and the distribution of Eotaxin genotypes. As shown, the wild-type genotype was most the common one, whereas the variant genotype was least common.

Table 1. Eotaxin ALA_{23}/THR_{23} genotype assignments and Hardy-Weinberg equilibrium

	Total	Genotype assignment						Hardy-Weinberg
		Wild-type		Heterozygous		Variant		
	N =	N =	%	N =	%	N =	%	*P* =
All subjects	233	166	72%	58	25%	9	4%	0.13
Gender:								
Male	130	86	66%	36	28%	8	6%	0.12
Female	102	80	78%	22	22%	1	1%	0.70
Ethnic origin:								
Caucasian	158	106	67%	46	29%	6	4%	0.72
African-American	40	32	80%	8	20%	0	0%	0.48
Asian	10	7	70%	2	20%	1	10%	0.24
Hispanic and other*	25	21	84%	2	8%	2	8%	0.002**

* Including subjects of unknown ethnicity
**P<0.05

The individual breakdown of population by race and gender indicated that all the subgroups studied were in Hardy Weinberg equilibrium except for the subgroup of Hispanics & others. However, the total number of subjects in this subgroup was relatively small and the subgroup itself is, by definition, not homogeneous.

DISCUSSION

Our study demonstrates that in the region on chromosome 17 containing the eotaxin gene (17q21.1 – q21.2), the EMGB is in Hardy-Weinberg equilibrium in all subpopulations except for "Hispanics and others". However, the size of the last subpopulation is small, and it is noteworthy that the addition of only three subjects of heterozygous genotype to this subgroup would put it back into Hardy-Weinberg equilibrium. Furthermore, this subpopulation is heterogeneous and accordingly might not be expected, *a priori*, to be in Hardy-Weinberg equilibrium. Overall, the EMGB is not under selection pressure at this locus and can therefore serve as a valid control population of healthy subjects for studies of genes near cytogenetic position 17q21.1 – q21.2.

Additional confirmation of the finding of Hardy-Weinberg equilibrium on the short arm of chromosome 17 was obtained in a recent study of the effect of angiotensin converting enzyme on physical performance among basic trainees (12). In this study, 147 samples from the EMGB were genotyped at the Angiotensin Converting Enzyme locus (located at cytogenetic position 17q23). As in this report, both the overall cohort and the major ethnic subgroups (Caucasian, African-American, Other) were found to be in Hardy-Weinberg equilibrium.

These results also demonstrate how an effective PCR-based strategy can be used to rapidly obtain SNP genotypes in the EMGB. PCR methods are relatively low-cost when compared to other methods such as gene sequencing. Furthermore, a high throughput of samples could be achieved by using 96 well plates, which allowed over 200 genotypes to be assigned over a three month work period.

REFERENCES

1. Bouchard, C., R. Malina, and L. Perusse. *Genetics of fitness and physical performance*. Champaign, IL: Human Kinetics, 1997.

2. Dalbey, R. E., M. O. Lively, S. Bron, and J. M. van Dijl. The chemistry and enzymology of the type I signal peptidases. *Protein Sci.* 6: 1129-1138, 1997.

3. Folz, R. J., S. F. Nothwehr, and J. I. Gordon. Substrate specificity of eukaryotic signal peptidase. Site-saturation mutagenesis at position -1 regulates cleavage between multiple sites in human pre (delta pro) apolipoprotein A-II. *J. Biol. Chem.* 263: 2070-2078, 1988.

4. Garcia-Zepeda, E. A., M. E. Rothenberg, S. Weremowicz, M. N. Sarafi, C. C. Morton, and A. D. Luster. Genomic organization, complete sequence, and chromosomal location of the gene for human eotaxin (SCYA11), an eosinophil-specific CC chemokine. *Genomics* 41: 471-476, 1997.

5. Griffiths, A. J. F., J. H. Miller, D. T. Suzuki, R. C. Lewontin, and W. M. Gelbart. *An introduction to genetic analysis*. New York, NY: W.H. Freeman and Company, 1996, p. 790-791 and 867.

6. Lilly, C. M., P. G. Woodruff, C. A. J. Camargo, H. Nakamura, J. M. Drazen, E. S. Nadel, and J. P. Hanrahan. Elevated plasma eotaxin levels in patients with acute asthma. *J. Allergy Clin. Immunol.* 104: 786-790, 1999.

7. Luster, A. D. Chemokines--chemotactic cytokines that mediate inflammation. *N. Engl. J. Med.* 338: 436-445, 1998.

8. Nakamura, H., A. D. Luster, T. Nakamura, K. H. In, L. A. Sonna, A. Deykin, E. Israel, J. M. Drazen, and C. M. Lilly. Variant eotaxin: its effects on the asthma phenotype. *J. Allergy Clin. Immunol.* 108: 946-953, 2001.

9. Newton, C. R., A. Graham, L. E. Heptinstall, S. J. Powell, C. Summers, N. Kalsheker, J. C. Smith, and A. F. Markham. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* 17: 2503-2516, 1989.

10. Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10: 1-6, 1997.

11. Sonna, L. A., K. C. Angel, M. Cullivan, and C. M. Lilly. *Environmental Medicine Genome Bank (EMGB).* Natick, MA: USARIEM Technical Note 1-99, 1999.

12. Sonna, L. A., M. A. Sharp, J. J. Knapik, M. Cullivan, K. C. Angel, J. F. Patton, and C. M. Lilly. Angiotensin-converting enzyme genotype and physical performance during U.S. Army basic training. *J. Appl. Physiol.* 91: 1355-1363, 2001.

13. Sonna, L. A., L. Zhao, K. C. Angel, M. Cullivan, and C. M. Lilly. *Environmental Medicine Genome Bank: Current Composition.* Natick, MA: USARIEM Technical Note 00-8, 2000.

APPENDIX: A THEORETICAL DEVIATION FROM HARDY-WEINBERG EQUILIBRIUM BASED ON POPULATION STRATIFICATION

Let us assume that the population under study is in fact stratified and consists of two subgroups of 1000 individuals, one with allele frequencies of p(0.8) and q(0.2), the other with allele frequencies of p(0.2) and q(0.8). Let us name the two alleles A and a. Let us also assume that the two subgroups are each individually in perfect Hardy-Weinberg equilibrium. Will the combined population be in Hardy-Weinberg equilibrium or not?

Assuming Hardy-Weinberg equilibrium, the number of individuals of each genotype in the first subgroup will be:

AA =	$(N)(p^2)$ =	$1000\ (0.8)^2$ =	640
Aa =	(N)(2pq) =	1000 (2)(0.8)(0.2)=	320
aa =	$(N)(q^2)$ =	$1000\ (0.2)^2$ =	40

Similarly, in the second subgroup, the number of individuals of each genotype will be AA = 40, Aa = 320, aa = 640. In the overall cohort, the number of individuals of each genotype will therefore be:

Genotype:	Subgroup 1 N=	Subgroup 2 N =	Mixed population observed N =	Mixed population genotype frequency
AA	640	40	680	0.34
Aa	320	320	640	0.32
aa	40	640	680	0.34
TOTAL	1000	1000	2000	1.00

Because each individual carries two alleles, one inherited from each parent, the total number of alleles in the mixed population is 2000 x 2 = 4000. Furthermore, in the mixed population there are 680 individuals with genotype AA (contributing 1360 'A' alleles to the total), 640 subjects with genotype Aa (contributing 640 'A' alleles and 640 'a' alleles) and 680 individuals of genotype aa (contributing another 1360 'a' alleles). Accordingly, the allele frequencies in the mixed population are:

$$f(A) = p = (1360 + 640) / (2 \times 2000) = 0.50$$

$$f(a) = q = (640 + 1360) / (2 \times 2000) = 0.50$$

From these allele frequencies, the expected distribution of genotypes in the mixed population based on the Hardy-Weinberg equilibrium assumption is:

AA =	$(N)(p^2)$ =	$2000\ (0.5)^2$ =	500
Aa =	(N)(2pq) =	2000 (2)(0.5)(0.5)=	1000
aa =	$(N)(q^2)$ =	$2000\ (0.5)^2$ =	500

A comparison of the observed and expected distributions yields the following, which allows computation of the Chi-squared statistic:

Genotype:	Observed N=	Expected N =		$(O-E)^2/E$
AA	680	500		64.8
Aa	640	1000		129.6
aa	680	500		64.8
TOTAL	2000	2000	χ^2 =	259.2

Because all 'non-A' alleles must be equal to 'a', there is only one degree of freedom in this distribution. The *P* value corresponding to this Chi-squared statistic is < 0.0001. Thus, even though the two subgroups were individually in perfect Hardy-Weinberg equilibrium, the distribution of genotypes in the population that resulted from the combining of the two would indeed deviate significantly from Hardy-Weinberg equilibrium.